**BIOST 2155 – Introductory Statistical Learning for Health Sciences – Fall 2024**

Credit Hours: 2 credits

Last updated: 8/18/2024

**Instructor:**

Lu Tang, PhD, Assistant Professor of Biostatistics

E-mail: [lutang@pitt.edu](mailto:lutang@pitt.edu)

Phone: (412) 383-0606

**Lecture Times and Location:**

Time: **Friday 9:00-10:50 AM** (8/26/2024–12/17/2024); no class on 11/29/2024 due to Thanksgiving Recess. See Course Schedule below for details. Calendar files are available for download on Canvas.

Location: A216 Public Health

Format: *In person only*

**Office Hours:** Fridays 11:00–11:40 AM in room A740, or by appointment; no office hours on 10/25/2024 and 11/29/2024, 12/13/2024. Calendar files are available for download on Canvas.

**Teaching Assistant:** NA

TA Office Hours: NA

**Course Description**

This 2-credit course is a graduate-level course to introduce basic concepts and methods for statistical learning with an emphasis on modern health science applications. The syllabus includes: linear regression with regularization, supervised machine learning, unsupervised clustering, and dimension reduction. The target audience will be second-year Biostatistics master students or early PhD students with interests in statistical learning techniques for health science data. Students will be trained with hands-on materials to understand the methods, implement the algorithms, and interpret results in real applications.

**Prerequisites**

BIOST 2039, 2049 and 2043. Students are expected to have programming experiences in R or in some low-level languages such as C, C++, Java and Fortran. Prerequisites can be met with other relevant courses/experience with consent from the instructor.

**Learning Objectives**

At the conclusion of this course, the students should be able to:

* Explain the motivation and insights behind statistical learning methods covered in class.
* Demonstrate basic theory proof for selected fundamental concepts and methods.
* Implement methods to real datasets in health sciences.
* Produce an analysis pipeline and interpret the results in a real application.

**Textbook**

Relevant lecture notes, slides and reading materials will be made available on Canvas. The following textbook is highly recommended and will be used for the majority of lectures.

* James, Witten, Hastie and Tibshirani. An Introduction of Statistical Learning: with Applications in R, Springer, 2013. (<https://www.statlearning.com/>)

The following two books are optional references. Free copies are available online.

* Hastie, Tibshirani, Friedman. The Element of Statistical Learning. Springer, 2009.
* Bishop. Pattern Recognition and Machine Learning. Springer New York, 2006.

**Canvas Instruction**

Canvas (<https://canvas.pitt.edu/courses/268942>) will serve as the main archive of lecture slides, lecture notes, homework assignments, and other materials. Announcements concerning course logistics will also be sent and posted on the webpage.

**Computing**

We will use [R](https://www.r-project.org/) as our main programming language. You can consider using [RStudio](https://posit.co/downloads/) and its R Markdown feature to generate homework reports. Template will be provided.

**Grading Scale and Student Performance Evaluation (Assessments and Weights)**

Course grades will be based on a weighted average of

* Homework assignments 60% (Three homework assignments; each 20%)
* Projects 40%:
  + Final project presentation (25%) (analysis of a real dataset or review of a methodological paper)
  + Kaggle prediction challenge (15%)

The cut-offs for computing letter grades will be: A, 100%-90%; B, 89%-80%; C, 79%-70%; D, 69%-60%; and F, <60%. Plus-minus grades will be assigned by dividing the respective intervals into thirds. Discussions of homework assignments among students are allowed. But each student must independently write and implement their own solution. Solutions will be graded on both correctness and clarity. Credits will be given for trying to identify the gaps in argument if complete solution cannot be derived. Cheating and plagiarism is strictly not allowed and may be reported to the university. See the University of Pittsburgh's Policy on Academic Integrity at <http://www.provost.pitt.edu/info/ai1.html>

Late homework will be discounted accordingly: 80% if 0-1 days past due, 50% if 1-2 days, 0% if later than 2 days. Exceptions can be made in case of sickness or emergency, under which the instructor must be informed in advance.

Homework

There will be 3 homework assignments (mixture of theory proof, simulation and real data application using R). Students will turn-in an electronic copy via Canvas. *Use white space and include clear comments to make code readable.*

Final projects

By 9/27, students should form final project groups (2-3 students per group) and determine a selected dataset or paper. In the meeting (12/6), each group will have 20 minutes to present their work and summarize their findings. Kaggle prediction challenge will run through the entire semester, closing on 12/13.

**Course Schedule**

|  |  |  |
| --- | --- | --- |
| Part I: Linear model and regularization | | |
| 8/30 | Introduction | ISLR, Ch2, 3 |
| 9/6 | Linear Model Selection and Regularization I (subset selection, ridge, lasso, elastic net) | ISLR, Ch6.1, 6.2 |
| 9/13 | Linear Model Selection and Regularization II (high-dimensionality, group lasso, fused lasso, GGM & graphical lasso)  HW1 distributed (due on 9/27) | ISLR, Ch6.4  ESL, Ch3.8.4, 18.4.2  ESL, Ch17.3 |
| 9/20 | Lab 1: R session |  |
| Part II: Unsupervised learning | | |
| 9/27 | Dimension reduction (PCA, MDS)  \* Deadline to form final project groups and select topic | ISLR, Ch6.3, 10.2  ESL, Ch14.8 |
| 10/4 | Clustering I (hierarchical clustering, K-means, ~~Gaussian mixture model~~) | ISLR, Ch10.3  Bishop, Ch9.1, 9.2 |
| 10/11 | Clustering II (estimate K, cluster evaluation)  HW2 distributed (due on 10/25) | Provided by instructor |
| 10/18 | Lab 2: R session |  |
| Part III: Supervised learning | | |
| 10/25 | Bayes classifier, LDA, QDA, KNN; | ISLR, Ch4 |
| 11/1 | Resampling methods  HW3 distributed (due on 12/13) | ISLR, Ch5 |
| 11/8 | Tree-Based Methods (CART, Bagging, Boosting, Random forest) | ISLR, Ch8 |
| 11/15 | Lab 3: R session |  |
| 11/22 | Support vector machines (SVM), Artificial Neural Network and Deep Learning | ISLR, Ch9  Bishop, Ch5 |
| 12/6 | Final project presentation | |
| 12/13 | NO CLASS |  |

ISLR: James, Witten, Hastie, Tibshirani. An Introduction to Statistical Learning: with Applications in R. Springer, 2013.

Bishop: Bishop. Pattern Recognition and Machine Learning. Springer New York, 2006.

ESL: Hastie, Tibshirani, Friedman. The Element of Statistical Learning. Springer, 2009.

**Disability Services Statement**

If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and [Disability Resources and Services](https://www.studentaffairs.pitt.edu/drs/) (DRS), 140 William Pitt Union, (412) 648-7890, [drsrecep@pitt.edu,](mailto:drsrecep@pitt.edu) (412) 228-5347 for P3 ASL users, as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.

**Academic Integrity Statement**

Students in this course will be expected to comply with the [University of Pittsburgh’s Policy on Academic Integrity](https://www.provost.pitt.edu/info/ai1.html). Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

To learn more about Academic Integrity, visit the [Academic Integrity Guide](http://pitt.libguides.com/academicintegrity/) for an overview of the topic. For hands- on practice, complete the [Understanding and Avoiding Plagiarism tutorial](http://pitt.libguides.com/academicintegrity/plagiarism).

**Sexual Misconduct, Required Reporting, and Title IX Statement**

If you are experiencing sexual assault, sexual harassment, domestic violence, and stalking, please report it to me and I will connect you to University resources to support you.

University faculty and staff members are required to report all instances of sexual misconduct, including harassment and sexual violence to the Office of Civil Rights and Title IX. When a report is made, individuals can expect to be contacted by the Title IX Office with information about support resources and options related to safety, accommodations, process, and policy. I encourage you to use the services and resources that may be most helpful to you.

As your professor, I am required to report any incidents of sexual misconduct that are directly reported to me. You can also report directly to Office of Civil Rights and Title IX: 412-648-7860  (M-F; 8:30am-5:00pm) or via the Pitt Concern Connection at: [Make A Report](https://app.convercent.com/en-US/LandingPage/2d6327d5-9fec-ea11-a974-000d3ab9f296?_=1612800567898)

An important exception to the reporting requirement exists for academic work. Disclosures about sexual misconduct that are shared as a relevant part of an academic project, classroom discussion, or course assignment, are not required to be disclosed to the University’s Title IX office.

If you wish to make a confidential report, Pitt encourages you to reach out to these resources:

* The University Counseling Center: 412-648-7930 (8:30 A.M. TO 5 P.M. M-F) and 412-648-7856 (AFTER BUSINESS HOURS)
* Pittsburgh Action Against Rape (community resource): 1-866-363-7273 (24/7)

If you have an immediate safety concern, please contact the University of Pittsburgh Police, 412-624-2121

Any form of sexual harassment or violence will not be excused or tolerated at the University of Pittsburgh.

For additional information, please visit the [full syllabus statement](https://www.diversity.pitt.edu/civil-rights-title-ix/make-report/responsible-employee-program-and-reporting#syllabus) on the Office of Diversity, Equity, and Inclusion webpage.

**Equity, Diversity, and Inclusion**The University of Pittsburgh does not tolerate any form of discrimination, harassment, or retaliation based on disability, race, color, religion, national origin, ancestry, genetic information, marital status, familial status, sex, age, sexual orientation, veteran status or gender identity or other factors as stated in the University’s Title IX policy. The University is committed to taking prompt action to end a hostile environment that interferes with the University’s mission. For more information about policies, procedures, and practices, visit the [Civil Rights & Title IX Compliance web page](https://www.diversity.pitt.edu/civil-rights-title-ix-compliance).

I ask that everyone in the class strive to help ensure that other members of this class can learn in a supportive and respectful environment. If there are instances of the aforementioned issues, please contact the Title IX Coordinator, by calling 412-648-7860, or e-mailing [titleixcoordinator@pitt.edu](mailto:titleixcoordinator@pitt.edu). Reports can also be [filed online](https://www.diversity.pitt.edu/civil-rights-title-ix-compliance/make-report/report-form). You may also choose to report this to a faculty/staff member; they are required to communicate this to the University’s Office of Diversity and Inclusion. If you wish to maintain complete confidentiality, you may also contact the University Counseling Center (412-648-7930).